CORRIGE DES EXERCICES : Exercices de révision

Exercice 8.1

$\mathcal{P}$={filles de 10 ans}, X= nombre de bonnes réponses au test des signes arithmétiques, variable quantitative normale de moyenne μ et d'écart-type σ, inconnus dans $\mathcal{P}$. On dispose d'un échantillon de X issu de la population $\mathcal{P}$ de taille n=24 sur lequel on estime μ par $\bar{x}$=10 et σ par s*=2,1 (estimation sans biais).

Test de comparaison d'une moyenne à une valeur théorique μ_0=11 :

test bilatéral de l'hypothèse nulle H_0 : $\mu=\mu_0=11$ contre l'hypothèse alternative H_1 : $\mu \neq \mu_0=11$ au risque α=5%. Puisque X est une variable normale, σ est inconnu et n=24<30, ce test est basé sur la statistique de test $T=\frac{(\overline{X}_n-\mu_0)\sqrt{n}}{S_n^*}$ qui suit une loi de Student $T_{n-1}=T_{23}$ sous H_0.

<u>règle de décision</u> :

on conserve H_0 (on rejette H_1) au seuil α=5% si la valeur observée de T, t appartient à la région d'acceptation du test bilatéral au seuil α=5% : $IA_{5\%}=[-t_{(1-\alpha)/2} ; t_{(1-\alpha)/2}]=[-t_{97,5\%} ; t_{97,5\%}[=[-2{,}069 ; 2{,}069]$ où $t_{97,5\%}=2{,}069$ est le quantile d'ordre $1-\frac{\alpha}{2}=0{,}975$ de la loi T_{23} (ν=23 et P=0,05) et ou

on rejette H_0 en faveur de H_1 (on accepte ou on valide H_1) au risque maximum α=5% sinon, c'est-à-dire si t n'appartient pas à $IA_{5\%}$.

La valeur observée de T : $t=\frac{(\bar{x}-\mu_0)\sqrt{n}}{s^*}=\frac{(10-11)\sqrt{24}}{2{,}1}=-2{,}333$.

<u>décision</u> : puisque t n'appartient pas à la région d'acceptation de H_0 on rejette donc H_0 en faveur de H_1 au risque α=5%.

➨ On peut accepter l'hypothèse que le nombre moyen de bonnes réponses chez les filles de 10 ans est différent de 11, au risque α=5%.

Exercice 8.2

$\mathcal{P}$={garçons de 12 à 15 ans}, X= temps pour loger 50 rondelles (en mn), variable quantitative de moyenne μ et d'écart-type σ, inconnus dans $\mathcal{P}$. On dispose d'un échantillon de X issu de la population $\mathcal{P}$ de taille n=37 sur lequel on estime μ par $\bar{x}$=1,42 mn et σ par s=0,24 mn (estimation biaisée).

Test de comparaison d'une moyenne à une valeur théorique μ_0=1,5 mn :

test unilatéral gauche de l'hypothèse nulle H_0 : $\mu=\mu_0=1{,}5$ contre l'hypothèse alternative H_1 : $\mu<\mu_0=1{,}5$ au risque α=1%. Puisque X est une variable quelconque, σ est inconnu et n=37>30, ce test est basé sur :

① la statistique de test $Z=\frac{(\overline{X}_n-\mu_0)\sqrt{n}}{S_n^*}$ qui suit approximativement une loi $\mathcal{N}(0,1)$ sous H_0.

L'hypothèse alternative étant unilatérale gauche, on rejettera l'hypothèse nulle H_0 pour les "petites" valeurs de Z, donc la région de rejet est à gauche du domaine de variation de la statistique de test Z.

<u>règle de décision</u> :

on rejette H_0 en faveur de H_1 (on accepte H_1 ou on valide H_1) au risque maximum α=1% si la valeur observée de Z, z appartient à la région de rejet du test unilatéral gauche au risque α=1% : $RC_{1\%}=]-\infty ; -z_{1-\alpha}[$ $RC_{1\%}=]-\infty ; -z_{0,99}[=]-\infty ; -2{,}325[$ où $-z_{0,99}=-2{,}325$ est le quantile d'ordre α=1% de la loi $\mathcal{N}(0,1)$

ou on conserve H_0 (on rejette H_1) au seuil α=1% sinon, c'est-à-dire si z n'appartient pas à $RC_{1\%}$.

La valeur observée de la statistique de test Z : $z=\frac{(\bar{x}-\mu_0)\sqrt{n}}{s^*}=\frac{(1{,}42-1{,}5)\sqrt{37}}{0{,}2433}=-2$ car l'estimation ponctuelle sans biais de σ : $s^*=\sqrt{\frac{37}{36}}0{,}24=0{,}2433$ mn

<u>décision</u> : puisque z n'appartient pas à la région de rejet de H_0 on ne rejette donc pas H_0 (on conserve H_0 ou on rejette H_1) au seuil α=1% et au risque β inconnu.

② la statistique de test $\overline{X}_n$ qui suit approximativement une loi $\mathcal{N}\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$ sous H_0.

<u>règle de décision</u> :
on rejette H_0 en faveur de H_1 (on accepte H_1 ou on valide H_1) au risque maximum α=1% si la valeur observée de $\overline{X}_n$, $\overline{x}$ appartient à la région de rejet du test unilatéral gauche au risque α=1% :

$RC_{1\%} = \left]+\infty\ ;\ \mu_0 - z_{1-\alpha}\frac{s^*}{\sqrt{n}}\right[= \left]-\infty\ ;\ 1{,}5-2{,}325\frac{0{,}2433}{\sqrt{37}}\right[= \left]-\infty\ ;\ 1{,}5-0{,}093\right[= \left]-\infty\ ;\ 1{,}407\right[$ car $s^* = \sqrt{\frac{37}{36}}0{,}24 = 0{,}2433$ mn et $z_{1-\alpha} = z_{0,99} = 2{,}325$ est le quantile d'ordre $1-\alpha$= 0,99 de la loi $\mathcal{N}(0,1)$ ou

on conserve H_0 (on rejette H_1) au seuil α=1% sinon, c'est-à-dire si z n'appartient pas à $RC_{1\%}$.

<u>décision</u> : puisque $\overline{x}$=1,42 n'appartient pas à $RC_{1\%}$ on ne rejette pas H_0 (on conserve H_0 ou on rejette H_1) au seuil α=1% et au risque β inconnu.

➨ On ne peut accepter l'hypothèse que le temps moyen des garçons de 12 à 15 ans est inférieur à 1,5 au seuil α=1% et au risque β inconnu.

Le risque minimum pour accepter H_1 (ou pour rejeter H_0) est le degré de signification α_{obs} : pour un test unilatéral gauche, c'est la probabilité d'obtenir sous H_0 une valeur de Z au moins aussi faible que celle observée z : $\alpha_{obs} = P(Z \le z) = P(Z \le -2) = F_Z(-2) = 1 - F_Z(2) = 1 - 0{,}9772 = 0{,}0228$. On pourra rejeter H_0 (accepter H_1) pour un risque $\alpha \ge \alpha_{obs} \approx 2{,}3\%$.

<u>Exercice 8.3</u>

$\mathcal{P}$={lancés d'un dé}
X= face obtenue définie sur E={ 1, 2, 3, 4, 5, 6} variable qualitative à 6 modalités.
On note :
p_1= proportion de 1, p_2= proportion de 2, p_3= proportion de 3, p_4= proportion de 4, p_5= proportion de 5 et p_6= proportion de 6, p_1, p_2, … p_6 étant inconnues dans $\mathcal{P}$.
L'hypothèse selon laquelle le dé est bien équilibré se traduit par le fait que les 6 proportions précédentes sont égales et s'écrit $H_0 : p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6$ (loi uniforme sur les 6 faces).
L'hypothèse selon laquelle le dé n'est pas équilibré (est pipé) se traduit par le fait qu'au moins une des proportions précédentes n'est pas égale à 1/6, il s'agit donc de l'hypothèse alternative H_1.

Test du khi-deux d'adéquation à une loi théorique au risque α=0,05 :

$\begin{cases} H_0 : \text{X suit la loi théorique : loi uniforme} \\ H_1 : \text{X ne suit pas la loi théorique} \end{cases}$ ou $\begin{cases} H_0 : p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6} \\ H_1 : \text{il existe i tel que } p_i \ne \frac{1}{6} \end{cases}$

Sur un échantillon de n=450 lancés, les effectifs observés n_i et les effectifs théoriques (attendus) sous H_0 $e_i = n \times \frac{1}{6} = 75$ sont donnés par :

X face	1	2	3	4	5	6	total n
effectif observé n_i	62	50	76	68	111	83	450
effectif attendu e_i	75	75	75	75	75	75	450

Sous H_0, la statistique de test Q^2 suit approximativement une loi du khi-deux à 5 ddl car n=60 ≥ 30 et tous les e_i sont supérieurs à 5. La région de rejet du test au risque α=0,05 est $RC_{0,05} =]q^2_{0,95}\ ;\ +\infty[=]11{,}07\ ;\ +\infty[$ et la région d'acceptation $IA_{5\%} = [0\ ;\ q^2_{0,95}] = [0\ ;\ 11{,}07]$ car $q^2_{0,95} = 11{,}070$ est le quantile d'ordre 0,95 de la loi χ^2_5.

La valeur observée de Q^2 :
$$q^2 = \frac{1}{75}\left[(-13)^2 + (-15)^2 + 1^2 + (-7)^2 + 36^2 + 8^2\right] = \frac{1804}{75} = 24{,}05$$
$q^2 \in RC_{0,05}$ donc on rejette H_0 au risque $\alpha = 0{,}05$.

➨ Il n'y a pas adéquation entre la loi de X et la loi théorique uniforme sur les 6 faces du dé au risque α=5% ; on ne peut pas conclure que le dé est équilibré au risque α=5%.

Exercice 8.4

$\mathcal{P}$={sujets privés de rêves}, X= score au test d'anxiété, variable quantitative de moyenne μ et d'écart-type σ, inconnus dans $\mathcal{P}$. On dispose d'un échantillon de X dans la population $\mathcal{P}$ de taille n=40 sur lequel on estime μ par $\overline{x}$=28,25 et σ par s=8,81 (estimation biaisée).

$\mathcal{P}_0$={sujets non privés de rêves} le score moyen au test d'anxiété est connu et vaut μ_0=26,5 dans $\mathcal{P}_0$.

Test de comparaison d'une moyenne à une valeur théorique μ_0=26,5 :

test unilatéral droit de l'hypothèse nulle H_0 : $\mu=\mu_0$=26,5 contre l'hypothèse alternative H_1 : $\mu>\mu_0$=25 au risque α=5%.
Puisque X est une variable quelconque, σ est inconnu et n=40 ≥ 30, ce test est basé sur la statistique de test

① la statistique de test $Z = \dfrac{\left(\overline{X}_n - \mu_0\right)\sqrt{n}}{S_n^*}$ qui suit approximativement une loi $\mathcal{N}(0,1)$ sous H_0.

La région de rejet du test unilatéral droit au risque α=5% : $RC_{5\%}$= $]z_{1-\alpha}$; $+\infty[=]z_{0,95}$; $+\infty[=]1,645; +\infty[$ car la valeur critique $z_{1-\alpha}=z_{0,95}$=1,645 est le quantile d'ordre 1−α=0,95 de la loi $\mathcal{N}(0,1)$.

La valeur observée de la statistique de test Z : $z = \dfrac{(\overline{x}-\mu_0)\sqrt{n}}{s^*} = \dfrac{(28,5-26,5)\sqrt{40}}{8,92}$ =1,24 car $s^* = \sqrt{\dfrac{40}{39}}\,8,81 = 8,92$. Puisque $z \notin RC_{5\%}$ on ne rejette pas H_0 (on conserve H_0) au seuil α=5% et au risque β inconnu.

② la statistique de test $\overline{X}_n$ qui suit approximativement une loi $\mathcal{N}\left(\mu_0 , \dfrac{\sigma}{\sqrt{n}}\right)$ sous H_0.

La région de rejet du test unilatéral droit au risque α=5% : $RC_{5\%} = \left]\ \mu_0 + z_{1-\alpha}\dfrac{s^*}{\sqrt{n}}\ ; +\infty\right[$

$RC_{5\%} = \left]\ 26,5+1,645\dfrac{8,92}{\sqrt{40}}\ ; +\infty\right[\ = \left]\ 26,5+2,32\ ; +\infty\right[= \left]\ 28,82\ ; +\infty\right[$ car $s^* = \sqrt{\dfrac{40}{39}}\,8,81 = 8,92$ et

$z_{1-\alpha}= z_{0,95}$= 1,645 est le quantile d'ordre 1−α= 0,95 de la loi $\mathcal{N}(0,1)$

Puisque $\overline{x}$=28,25 n'appartient pas à $RC_{5\%}$ on ne rejette donc pas H_0 (on conserve H_0 ou on rejette H_1) au seuil α=% et au risque β inconnu.

➡ On ne peut accepter l'hypothèse que la privation de rêves augmente le niveau d'anxiété au seuil α=5% et au risque β inconnu.

Le risque minimum pour accepter H_1 (ou pour rejeter H_0) est le degré de signification α_{obs} : pour un test unilatéral droit, c'est la probabilité d'obtenir sous H_0 une valeur de Z au moins aussi élevée que celle observée z, c'est à dire que $\alpha_{obs} = P(Z \ge z) = P(Z \ge 1,24) = 1 - F_Z(1,24) = 1 - 0,5948 = 0,4052$ donc α_{obs} ≈40,5%. On pourrait rejeter H_0 (accepter H_1) si le risque maximum $\alpha \ge \alpha_{obs}$ ≈40,5%.

Exercice 8.5

$\mathcal{P}$ ={personnes atteintes de schizophrénie}, X= volume de l'hippocampe gauche (en cm^3), variable quantitative normale de moyenne μ et d'écart-type σ, inconnus dans $\mathcal{P}$. On dispose d'un échantillon de X issu de la population $\mathcal{P}$ de taille n=15 sur lequel on observe : Σx_i = 23,4 et Σx_i^2 = 37,78.

Test de comparaison d'une moyenne à une valeur théorique μ_0=1,7 :

test unilatéral gauche de l'hypothèse nulle H_0 : $\mu=\mu_0$=1,7 contre l'hypothèse alternative H_1 : $\mu < \mu_0$=1,7 au risque α=1%. Puisque X est une variable normale, σ est inconnu et n=15<30, ce test est basé sur la statistique de test $T=\dfrac{\left(\overline{X}_n - \mu_0\right)\sqrt{n}}{S_n^*}$ qui suit une loi de Student $T_{n-1}=T_{14}$ sous H_0.

<u>règle de décision</u> :
on rejette H_0 en faveur de H_1 (on accepte ou on valide H_1) au risque maximum α=1% si la valeur observée de T, t appartient à la région de rejet du test unilatéral gauche au seuil α=1% : $RC_{1\%} =]-\infty\ ; -t_{1-\alpha}[=]-\infty\ ; -t_{99\%}[$
$RC_{1\%} =]-\infty\ ; -2,624[$ où $t_{99\%}$ = 2,624 est le quantile d'ordre 1−α=0,99 de la loi T_{14} (ν=14 et P=0,02) et ou
on conserve H_0 (on rejette H_1) au seuil α=1% sinon, c'est-à-dire si t n'appartient pas à $RC_{1\%}$.

La valeur observée de T : $t = \frac{(\overline{x} - \mu_0)\sqrt{n}}{s^*} = \frac{(1{,}56 - 1{,}7)\sqrt{15}}{0{,}3} = -1{,}796$ car $\overline{x} = \frac{23{,}4}{15} = 1{,}56$

$s^2{}^* = \frac{37{,}78 - (15 \times 1{,}56^2)}{14} = 0{,}0911$ et s*=0,3

(ou $s^2 = \frac{37{,}78}{15} - (1{,}56^2) = 0{,}085$ $s = \sqrt{0{,}085} = 0{,}292$ et $s^* = \sqrt{\frac{15}{14}} \times 0{,}292 \approx 0{,}3$)

<u>décision</u> : puisque t n'appartient pas à la région de rejet de H_0 on ne rejette pas H_0 (on conserve H_0 ou on rejette H_1) au seuil α=5% et au risque d'erreur de 2^d^ espèce β inconnu.

➨ On ne peut pas accepter l'hypothèse que le volume moyen de l'hippocampe gauche des sujets atteints de schizophrénie est inférieur à 1,7 cm^3, au seuil α=5% et au risque d'erreur de 2^d espèce β inconnu.

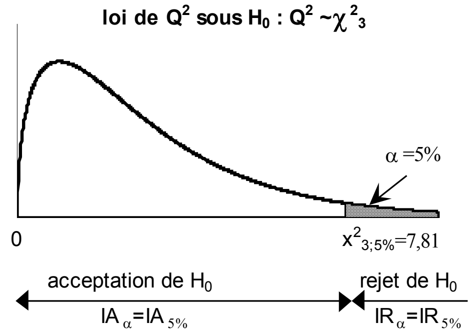
<u>Exercice 8.6</u>

$\mathcal{P}$={patients d'un hôpital psychiatrique}, X=saison définie sur E={printemps, été, automne, hiver} variable qualitative à 4 modalités, Y= réaction à un certain médicament définie sur F={oui, non} variable qualitative à 2 modalités.

Test du khi-deux d'indépendance : $\begin{cases} H_0: \text{X et Y indépendantes} \\ H_1: \text{X et Y liées} \end{cases}$ au risque α=5%.

Sur deux échantillons appariés de X et de Y de taille n=490, les effectifs observés n_{ij} et les effectifs théoriques (attendus) sous H_0 e_{ij} notés entre parenthèses :

X saison	Y réaction oui		non		total
printemps	55	(54,9)	64	(64,1)	119
été	59	(54,9)	60	(64,1)	119
automne	52	(53)	63	(62)	115
hiver	60	(63,2)	77	(73,8)	137
Total	226		264		**n=490**



Sous H_0, la statistique de test Q^2 suit approximativement une loi du khi-deux à 3 ddl, car n=490 ≥ 30 et tous les e_{ij} sont supérieurs à 5. La région de rejet du test au risque 5% est $RC_{5\%} =]q^2_{95\%}\ ; +\infty[\ =]7{,}815\ ; +\infty[$ et la région d'acceptation $IA_{5\%} = [0\ ;\ q^2_{95\%}] = [0\ ; 7{,}815]$ car $q^2_{95\%} = 7{,}815$ est le quantile d'ordre 0,95 de la loi χ^2_3.

La valeur observée de Q^2 vaut :

$$q^2 = \frac{(55-54{,}9)^2}{54{,}9} + \frac{(64-64{,}1)^2}{64{,}1} + \frac{(59-54{,}9)^2}{54{,}9} + \frac{(60-64{,}1)^2}{64{,}1} + \frac{(52-53)^2}{53} + \frac{(63-62)^2}{62} + \frac{(60-63{,}2)^2}{63{,}2} + \frac{(77-73{,}8)^2}{73{,}8}$$

$$= \frac{0{,}1^2}{54{,}9} + \frac{(-0{,}1)^2}{64{,}1} + \frac{4{,}1^2}{54{,}9} + \frac{(-4{,}1)^2}{64{,}1} + \frac{(-1)^2}{53} + \frac{1^2}{62} + \frac{(-3{,}2)^2}{63{,}2} + \frac{3{,}2^2}{73{,}8}$$

$$= 0{,}1^2\left[\frac{1}{54{,}9} + \frac{1}{64{,}1}\right] + 4{,}1^2\left[\frac{1}{54{,}9} + \frac{1}{64{,}1}\right] + 1^2\left[\frac{1}{53} + \frac{1}{62}\right] + 3{,}2^2\left[\frac{1}{63{,}2} + \frac{1}{73{,}8}\right] = 0{,}90455$$

q^2= 0,905 $\notin RC_{5\%}$ donc on ne rejette pas H_0 au seuil α=5%.

➨ Il n'existe pas de lien entre la saison et la réaction au seuil α=5% et au risque de 2ème espèce β inconnu.

<u>Exercice 8.7</u>

$\mathcal{P}$={ouvriers travaillant en usine}, X= nombre d'accidents subis dans l'année définie sur E={0, 1, 2, 3 ou plus} variable qualitative à 4 modalités.

On note p_1= proportion d'ouvriers n'ayant subi aucun accident, p_2= proportion d'ouvriers ayant subi 1 accident, p_3= proportion d'ouvriers ayant subi 2 accidents, p_4= proportion d'ouvriers ayant subi 3 accidents ou plus, p_1, p_2, … p_4 étant inconnues dans $\mathcal{P}$.

L'hypothèse émise selon laquelle le nombre d'accidents subis dans l'année suit la loi théorique s'écrit H_0 : p_1=0,57, p_2=0,32 p_3= 0,08 et p_4= 0,03

Test du khi-deux d'adéquation à une loi théorique au risque α=5% :

$\begin{cases} H_0 : \text{X suit la loi théorique} \\ H_1 : \text{X ne suit pas la loi théorique} \end{cases}$ ou $\begin{cases} H_0 : p_1 = 0{,}57 \quad p_2 = 0{,}32 \quad p_3 = 0{,}08 \quad p_4 = 0{,}03 \\ H_1 : \text{il existe i tel que } p_i \text{ différent de la loi théorique} \end{cases}$

Sur un échantillon de n=220 ouvriers, les effectifs observés n_i et les effectifs théoriques (attendus) e_i sous H_0 sont donnés par :

X nombre d'accidents subis	0	1	2	3 ou plus	total n
effectif observé n_i	154	50	15	1	220
effectif attendu e_i	220×0,57=125,4	220×0,57=70,4	220×0,57=17,6	220×0,57=6,6	220

Sous H_0, la statistique de test Q^2 suit approximativement une loi du khi-deux à 3 ddl car n=220 ≥ 30 et tous les e_i sont supérieurs à 5. La région de rejet du test au risque α=5% est $RC_{5\%} =]\, q^2_{95\%}\,;\, +\infty[\, =]7{,}815\,;\, +\infty[$ et la région d'acceptation $IA_{5\%} = [0\,;\, q^2_{95\%}\,] = [0\,;\, 7{,}815]$ car $q^2_{95\%} = 7{,}815$ est le quantile d'ordre 0,95 de la loi χ^2_3.

La valeur observée de Q^2 :

$$q^2 = \frac{(154-125{,}4)^2}{125{,}4}+\frac{(50-70{,}4)^2}{70{,}4}+\frac{(15-17{,}6)^2}{17{,}6}+\frac{(1-6{,}6)^2}{6{,}6}=\frac{(28{,}6)^2}{125{,}4}+\frac{(-20{,}4)^2}{70{,}4}+\frac{(-2{,}6)^2}{17{,}6}+\frac{(-5{,}6)^2}{6{,}6}$$

$q^2 = 17{,}570 \in RC_{1\%}$ donc on rejette H_0 en faveur de H_1 au risque α=5%.

➨ On ne peut pas accepter l'hypothèse émise : il n'y a pas adéquation entre la loi de X et la loi théorique sur les nombres d'accidents subis par an par les ouvriers travaillant en usine au risque α=5%.

Exercice 8.8

$\mathcal{P}_1$={enfant de 7 ans de mère alcoolique chronique durant la grossesse}
$\mathcal{P}_2$={enfant de 7 ans de mère n'ayant eu aucune tendance à l'alcoolisme durant la grossesse}
X = score de QI de l'enfant, variable quantitative
X_1= score de QI de l'enfant dans $\mathcal{P}_1$ de moyenne μ_1 et d'écart-type σ_1 inconnus dans $\mathcal{P}_1$
X_2= score de QI de l'enfant dans $\mathcal{P}_2$ de moyenne μ_2 et d'écart-type σ_2 inconnus dans $\mathcal{P}_2$
On observe :
un échantillon de X_1 issu de $\mathcal{P}_1$ de taille $n_1 = 6$ pour lequel

le score moyen μ_1 est estimé par $\overline{x}_1 = 78$ et l'écart-type du score σ_1 est estimé par s_1=8,18 (estimation biaisée)

un échantillon de X_2 issu de $\mathcal{P}_2$ de taille n_2=12 pour lequel

le score moyen μ_2 est estimé par $\overline{x}_2 = 99$ et l'écart-type du score σ_2 est estimé par s_2=10,13 (estimation biaisée)

L'hypothèse que l'alcool contrarie le développement cérébral prénatal se traduit par le fait que le score de QI moyen des enfants de mère alcoolique chronique durant la grossesse μ_1 est inférieur à celui des enfants de mère n'ayant eu aucune tendance à l'alcoolisme durant la grossesse μ_2 : $\mu_1 < \mu_2$. Tester l'hypothèse émise revient donc à comparer les deux moyennes μ_1 et μ_2 à partir de deux échantillons indépendants en utilisant le test unilatéral de l'hypothèse nulle $H_0 : \mu_1 = \mu_2$ contre l'hypothèse alternative $H_1 : \mu_1 < \mu_2$ au risque α=5%.

Test de comparaison de deux moyennes μ_1 et μ_2 à partir de deux échantillons indépendants :

test de l'hypothèse nulle $H_0 : \mu_1 = \mu_2$ contre l'hypothèse alternative $H_1 : \mu_1 < \mu_2$ au risque unilatéral α=5%.

Les variables X_1 et X_2 étant normales dans $\mathcal{P}_1$ et $\mathcal{P}_2$ et de variances inconnues égales (n_1=6 < 30 et n_2=12 < 30) ce test est basé sur :

① la statistique de test de Student $T = \dfrac{\overline{X}_1 - \overline{X}_2}{S^* \sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$ qui suit une loi de Student $T_{n_1+n_2-2}=T_{16}$ sous H_0.

Sous l'hypothèse alternative unilatérale $H_1 : \mu_1 < \mu_2$ les valeurs de $\overline{X}_1$ ont tendance à être inférieures à celles de $\overline{X}_2$ donc les valeurs de T ont tendance à être négatives, c'est-à-dire que la région de rejet du test est à gauche du domaine de variation de T.

La région de rejet du test unilatéral au risque α=5% : $RC_{5\%}=]-\infty\,;\,-t_{0,95}[=]-\infty\,;\,-1{,}746[$ où $t_{0,95}=1{,}746$ est le quantile d'ordre 1−α=0,95 de la loi T_{16} (ν=16 et P=0,10).

L'écart-type observé biaisé de X_1 dans $\mathcal{P}_1$: s_1=8,18 et l'écart-type observé biaisé de X_2 dans $\mathcal{P}_2$: s_2=10,13

La variance commune σ^2 de X_1 dans $\mathcal{P}_1$ et de X_2 dans $\mathcal{P}_2$ est estimée par la variance observée sans biais :

$$s^{*2} = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2} = \frac{6\times 8{,}18^2 + 12\times 10{,}13^2}{6+12-2} = 102{,}05 \text{ d'où } s^* \approx 10{,}1$$

La valeur observée de la statistique de test T : $t = \dfrac{\overline{x}_1 - \overline{x}_2}{s^*\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} = \dfrac{78-99}{10{,}1\sqrt{\dfrac{1}{6}+\dfrac{1}{12}}} = \dfrac{-21}{5{,}05} = -4{,}158$

règle de décision

- on rejette H_0 en faveur de H_1 au risque α=5% si la valeur observée de T, t appartient à $RC_{5\%} =]-\infty\ ;\ -1{,}746[$ la région de rejet du test unilatéral au risque α=5% et
- on conserve H_0 (on rejette H_1) au seuil α=5% sinon, c'est-à-dire si t n'appartient pas à $RC_{5\%}$.

décision

La valeur observée t appartient à la région de rejet de H_0 : on rejette donc H_0 en faveur de H_1 au risque α=5%.

② la statistique de test de Student $T = \dfrac{\overline{X}_2 - \overline{X}_1}{S^* \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ qui suit une loi de Student $T_{n_1+n_2-2} = T_{16}$ sous H_0.

Sous l'hypothèse alternative unilatérale $H_1 : \mu_2 > \mu_1$ les valeurs de $\overline{X}_2$ ont tendance à être supérieures à celles de $\overline{X}_1$ donc les valeurs de T ont tendance à être positives, c'est-à-dire que la région de rejet du test est à droite du domaine de variation de T.

La région de rejet du test unilatéral au risque α=5% : $RC_{5\%} =]\ t_{0,95}\ ;\ +\infty[=]\ 1{,}746;\ +\infty[$ où $t_{0,95}$ =1,746 est le quantile d'ordre $1-\alpha$=0,95 de la loi T_{16} (ν=16 et P=0,10).

La valeur observée de la statistique de test T : $t = \dfrac{\overline{x}_2 - \overline{x}_1}{s^* \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \dfrac{99-78}{10{,}1\sqrt{\dfrac{1}{6} + \dfrac{1}{12}}} = \dfrac{21}{5{,}05} = 4{,}158$ car $s^* \approx 10{,}1$.

règle de décision

- on rejette H_0 en faveur de H_1 au risque α=5% si la valeur observée de T, t appartient à $RC_{5\%} =]1{,}746;\ +\infty[$ la région de rejet du test unilatéral au risque α=5% et
- on conserve H_0 (on rejette H_1) au seuil α=5% sinon, c'est-à-dire si t n'appartient pas à $RC_{5\%}$.

décision

La valeur observée t appartient à la région de rejet de H_0 : on rejette donc H_0 en faveur de H_1 au risque α=5%.

➨ Le score de QI moyen des enfants de mère alcoolique chronique durant la grossesse est inférieur à celui des enfants de mère n'ayant eu aucune tendance à l'alcoolisme durant la grossesse au risque α=5% , ce qui confirme l'hypothèse que l'alcool contrarie le développement cérébral prénatal, au risque α=5%.

Exercice 8.9

$\mathcal{P}$={personnes}, X=opinion sur l'avortement définie sur E={pour, indifférent, contre} variable qualitative à 3 modalités, Y= nombre d'années de scolarité définie sur F={moins de 8 ans, entre 8 et 12 ans, plus de 12 ans} variable qualitative à 3 modalités.

1) **Test du khi-deux d'indépendance** : $\begin{cases} H_0 : \text{ X et Y indépendantes} \\ H_1 : \text{ X et Y liées} \end{cases}$ au risque α=5%.

Sur deux échantillons appariés de X et de Y de taille n=776, les effectifs observés n_{ij} et les effectifs théoriques (attendus) sous H_0 e_{ij} notés entre parenthèses :

X scolarité	Y opinion sur l'avortement pour	indifférent	contre	total
moins de 8 ans	31 (45,1)	23 (21,4)	56 (43,5)	110
entre 8 et 12 ans	171 (179,1)	89 (85,0)	177 (172,9)	437
plus de 12 ans	116 (93,8)	39 (44,6)	74 (90,6)	229
total	318	151	307	**n=776**

Sous H_0, la statistique de test Q^2 suit approximativement une loi du khi-deux à 4 ddl car n=776 ≥ 30 et tous les e_{ij} sont supérieurs à 5. La région de rejet du test au risque α=0,05 est $RC_{5\%} =]\ q^2_{0,95}\ ;+\infty[=]\ 9{,}488\ ;+\infty\ [$ et la région d'acceptation $IA_{5\%} = [\ 0\ ;\ 9{,}488]$ car $q^2_{0,95} = 9{,}488$ est le quantile d'ordre 0,95 de la loi χ^2_4.

La valeur observée de Q^2 vaut :

$$q^2 = \frac{(31-45{,}1)^2}{45{,}1} + \frac{(23-21{,}4)^2}{21{,}4} + \frac{(56-43{,}5)^2}{43{,}5} + \frac{(171-179{,}1)^2}{179{,}1} + \frac{(89-85)^2}{85} + \frac{(177-172{,}9)^2}{172{,}9}$$
$$+ \frac{(116-93{,}8)^2}{93{,}8} + \frac{(39-44{,}6)^2}{44{,}6} + \frac{(74-90{,}6)^2}{90{,}6} = 17{,}708$$

q^2= 17,708 ∈ $RC_{5\%}$ donc on rejette H_0 au risque α=5%.

➨ Il existe un lien entre l'opinion sur l'avortement et la durée de la scolarité au risque α=5%.

2) $\mathcal{P}$ ={personnes dont la durée de scolarité est inférieure à 8 ans}

X= rejet de l'avortement (opinion contre) définie sur E={oui, non} variable qualitative à 2 modalités avec p= proportion de personnes qui rejettent l'avortement, inconnue dans $\mathcal{P}$.

On dispose d'un échantillon de X issu de la population $\mathcal{P}$, de taille n=31+23+56=110 sur lequel on estime p par $f = \frac{56}{110} = 0{,}509$.

L'hypothèse que le rejet est majoritaire pour ces individus correspond au fait que dans la population $\mathcal{P}$, la proportion de ceux qui rejettent l'avortement p est supérieure à p_0=50%.

Test de comparaison d'une proportion à la proportion théorique p_0=50%=0,5 :
$H_0 : p = p_0 = 0{,}5$ contre $H_1 : p > p_0 = 0{,}5$ alternative unilatérale droite, au risque α=5%.
Sous H_0, puisque $n=110 \geq 30$, $np_0 = n(1-p_0) = 110\times 0{,}5 = 55 \geq 5$, ce test est basé sur la statistique de test :

① $Z = \frac{(F_n - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} = \frac{(F_n - 0{,}5)\sqrt{110}}{\sqrt{0{,}5(1-0{,}5)}}$ qui suit approximativement une loi $\mathcal{N}(0,1)$ sous H_0

Sous l'hypothèse alternative unilatérale $H_1 : p > p_0=0{,}5$ les valeurs de F_n ont tendance à être supérieures à p_0=0,5 donc les valeurs de Z ont tendance à être positives, c'est-à-dire que la région de rejet du test est à droite du domaine de variation de Z.
La région de rejet du test unilatéral droit au risque α=5% : $RC_{5\%}=]z_{1-\alpha}\ ;+\infty[=]z_{0,95}\ ;+\infty[=]1{,}645;+\infty[$ car la valeur critique $z_{1-\alpha}=z_{0,95}=1{,}645$ est le quantile d'ordre $1-\alpha=0{,}95$ de la loi $\mathcal{N}(0,1)$.

<u>règle de décision</u>

- on rejette H_0 en faveur de H_1 au risque α=5% si la valeur observée de Z, z appartient à la région de rejet du test unilatéral droit au risque α=5% : $RC_{5\%}=]z_{1-\alpha}\ ;+\infty[=]z_{0,95}\ ;+\infty[=]1{,}645;+\infty[$ et
- on conserve H_0 (on rejette H_1) au seuil α=5% sinon, c'est-à-dire si z n'appartient pas à $RC_{5\%}$.

<u>décision</u>

La valeur observée de la statistique de test Z : $z = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} = \frac{(0{,}509 - 0{,}5)\sqrt{110}}{\sqrt{0{,}5\times 0{,}5}} = \frac{0{,}095}{0{,}5} = 0{,}191$

Puisque z n'appartient pas à la région de rejet de H_0 on conserve donc H_0 (on rejette H_1) au seuil α=5%.

② F_n qui suit approximativement une loi $\mathcal{N}\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$ sous H_0

Sous l'hypothèse alternative unilatérale $H_1 : p > p_0=0{,}5$ les valeurs de F_n ont tendance à être supérieures à p_0=0,5, c'est-à-dire que la région de rejet du test est à droite de p_0=0,5.

La région de rejet du test unilatéral droit au risque α=5% : $RC_\alpha = \left]p_0 + z_{1-\alpha}\sqrt{\frac{p_0(1-p_0)}{n}};+\infty\right[$ d'où

$RC_{5\%}$ =] 0,5 + $z_{0,95}$×0,0477 ; +∞ [=] 0,5 + 1,645×0,0477 ; +∞ [=] 0,5 + 0,0784 ; +∞ [=] 0,578 ; +∞ [
où $z_{1-\alpha} = z_{0,95}=1{,}645$ est le quantile d'ordre $1-\alpha=0{,}95$ de la loi $\mathcal{N}(0,1)$.

<u>règle de décision</u>

- on rejette H_0 (on accepte H_1 ou on valide H_1) au risque α=5% si la valeur observée de F_n, f appartient à la région de rejet du test unilatéral droit au risque α=5% : $RC_{5\%}$ =]0,578 ; +∞[et
- on conserve H_0 (on rejette H_1) au seuil α=5% sinon, c'est-à-dire si z n'appartient pas à $RC_{5\%}$.

<u>décision</u>

La valeur observée de la statistique de test F_n : f = 0,509
Puisque f n'appartient pas à la région de rejet de H_0 on conserve donc H_0 (on rejette H_1) au seuil α=5%.

➨ On ne peut pas accepter l'hypothèse que le rejet est majoritaire (plus de 50%) chez les personnes dont la scolarité est inférieure à 8 ans, au seuil α=5% et au risque de seconde espèce β inconnu.

Le risque minimum pour accepter H_1 (ou pour rejeter H_0) est le degré de signification α_{obs} :
c'est la probabilité d'obtenir sous H_0 une valeur de Z au moins aussi élevée que celle observée z, c'est à dire que $\alpha_{obs} = P(Z \geq z) = P(Z \geq 0{,}189) = 1 - F_Z(0{,}189) \approx 1 - F_Z(0{,}19) = 1 - 0{,}5753 = 0{,}4247$.
On accepterait H_1 (rejetterait H_0) pour un risque maximum $\alpha \geq \alpha_{obs} \approx 42{,}5\%$ risque minimum.
($\alpha_{obs} = P_{H_0}(F_n \geq f) = P_{H_0}(F_n \geq 0{,}509) = P(Z \geq z) = P(Z \geq 0{,}189) \approx 1-F_Z(0{,}19)=0{,}4247$).

Exercice 8.10

$\mathcal{P}_1$={patients ayant un médecin traitant choisi}, $\mathcal{P}_2$={patients ayant un médecin traitant imposé},
X=score au test mesurant le niveau de satisfaction, variable quantitative
notée X_1 dans $\mathcal{P}_1$ de moyenne μ_1 et d'écart-type σ_1, inconnus dans $\mathcal{P}_1$ et
notée X_2 dans $\mathcal{P}_2$ de moyenne μ_2 et d'écart-type σ_2, inconnus dans $\mathcal{P}_2$.

On dispose d'un échantillon de X_1 issu de $\mathcal{P}_1$ de taille n_1=35 et d'un échantillon de X_2 issu de $\mathcal{P}_2$ de taille n_2=35, indépendants sur lesquels on estime μ_1 par $\overline{x}_1$=65,07 et μ_2 par $\overline{x}_2$=56,47.

Test de comparaison de deux moyennes à partir de deux échantillons indépendants :
test unilatéral de l'hypothèse nulle H_0 : $\mu_1=\mu_2$ contre l'hypothèse alternative H_1 : $\mu_1>\mu_2$ au risque α=5%. Puisque les tailles des échantillons n_1=35>30 et n_2=35>30 le test est basé sur la statistique de test $Z=\dfrac{\overline{X}_1-\overline{X}_2}{\sqrt{\dfrac{S_1^{*2}}{n_1}+\dfrac{S_2^{*2}}{n_2}}}$ qui suit approximativement une loi $\mathcal{N}(0,1)$ sous H_0. La région de rejet du test unilatéral au risque α=5% est $RC_{5\%}$ =]1,645 ; +∞] où la valeur critique $z_{1-\alpha}=z_{0,95}$= 1,645 est le quantile d'ordre α=0,95 de la loi $\mathcal{N}(0,1)$.

La variance observée sans biais dans $\mathcal{P}_1$: $s_1^{*2}=\dfrac{n_1}{n_1-1}s_1^2=\dfrac{35}{34}15{,}22^2=238{,}46$ (s_2*=15,44)

la variance observée sans biais dans $\mathcal{P}_2$: $s_2^{*2}=\dfrac{n_2}{n_2-1}s_2^2=\dfrac{35}{34}14{,}58^2=218{,}83$ (s_1*=14,79)

la valeur observée de la statistique de test Z : $z=\dfrac{\overline{x}_1-\overline{x}_2}{\sqrt{\dfrac{s_1^{*2}}{n_1}+\dfrac{s_2^{*2}}{n_2}}}=\dfrac{65{,}07-56{,}47}{\sqrt{\dfrac{218{,}83}{35}+\dfrac{238{,}46}{35}}}=2{,}387$ appartient à la région de rejet de H_0 : on rejette H_0 en faveur de H_1 au risque α=5%.

➨ On peut accepter l'hypothèse que la possibilité de choisir son médecin a une influence bénéfique sur le niveau de satisfaction, au risque α=5%.

Exercice 8.13

$\mathcal{P}_1$={ouvriers de l'entreprise}, X_1= opinion des ouvriers vis à vis de la réforme
$\mathcal{P}_2$={cadres moyens de l'entreprise}, X_2= opinion des cadres moyens vis à vis de la réforme
$\mathcal{P}_3$={cadres supérieurs de l'entreprise}, X_3= opinion des cadres supérieurs vis à vis de la réforme
X_1 , X_2 et X_3 variables qualitatives à 2 modalités définies sur E={favorable, opposé}.

Test du khi-deux d'homogénéité sur 3 populations avec une variable à 2 modalités c'est à dire un test de comparaison de trois proportions sur 3 échantillons indépendants (procédure bilatérale) :

$$\begin{cases} H_0 : p_1 = p_2 = p_3 \\ H_1 : p_1 \neq p_2 \text{ ou } p_1 \neq p_3 \text{ ou } p_2 \neq p_3 \end{cases} \quad \text{où} \quad \begin{bmatrix} p_1 = \text{proportion d'opinion favorable chez les ouvriers} \\ p_2 = \text{proportion d'opinion favorable chez les cadres moyens} \\ p_3 = \text{proportion d'opinion favorable chez les cadres supérieurs} \end{bmatrix}$$

Sur trois échantillons indépendants de X_1 , X_2 et X_3 de tailles n_1=285, n_2=75 et n_3=40, les effectifs observés n_{ij} et les effectifs théoriques (attendus) sous H_0 e_{ij} notés entre parenthèses :

X opinion	ouvriers		cadres moyens		cadres supérieurs		total
favorable	184	(188,1)	49	(49,5)	31	(26,4)	264
opposé	101	(96,9)	26	(25,5)	9	(13,6)	136
total	**n_1=285**		**n_2=75**		**n_3=40**		n=400

Sous H_0, la statistique de test Q^2 suit approximativement une loi du khi-deux à 2 ddl car $n=n_1+n_2+n_3=400 \geq 30$ et tous les e_{ij} sont supérieurs à 5. La région de rejet du test au risque α=1% est $RC_{1\%}$=]$q^2_{99\%}$; +∞[=]9,210 ; +∞[car $q^2_{99\%}$ est le quantile d'ordre 0,99 de la loi χ^2_2, et la valeur observée de Q^2 :

$$q^2 = \frac{(184-188{,}1)^2}{188{,}1} + \frac{(49-49{,}5)^2}{49{,}5} + \frac{(31-26{,}4)^2}{26{,}4} + \frac{(101-96{,}9)^2}{96{,}9} + \frac{(26-25{,}5)^2}{25{,}5} + \frac{(9-13{,}6)^2}{13{,}6}$$

$$q^2 = 4{,}1^2\left[\frac{1}{188{,}1}+\frac{1}{96{,}9}\right] + 0{,}5^2\left[\frac{1}{49{,}5}+\frac{1}{25{,}5}\right] + 4{,}6^2\left[\frac{1}{26{,}4}+\frac{1}{13{,}6}\right] = 0{,}263 + 0{,}015 + 2{,}357 = 2{,}635$$

$q^2 = 2{,}635 \notin RC_{10\%}$ donc on ne rejette pas H_0 au seuil α=1% et au risque de 2ème espèce β inconnu.

➨ Dans l'entreprise, l'opinion vis à vis de la réforme ne diffère pas selon le type d'emploi, au seuil α=1% et au risque de 2ème espèce β. Les proportions d'opinion favorable à la réforme ne sont pas différentes selon le type d'emploi, au seuil α=10% et au risque de 2ème espèce β.